Astronomical Data Analysis Software and Systems XVIII
ASP Conference Series, Vol. 411, *© 2009*
D. Bohlender, D. Durand and P. Dowler, eds.

High-Dimensional Data Reduction, Image Inpainting and their Astronomical Applications

Meyer Pesenson, *Isaac Pesenson, Sean Carey, Bruce McCollum, and William Roby

*SSC, California Institute of Technology, *Temple University, U.S.A.*

Abstract. Technological advances are revolutionizing multispectral astrophysics as well as the detection and study of transient sources. This new era of multitemporal and multispectral data sets demands new ways of data representation, processing and management thus making data dimension reduction instrumental in efficient data organization, retrieval, analysis and information visualization. Other astrophysical applications of data dimension reduction which require new paradigms of data analysis include knowledge discovery, cluster analysis, feature extraction and object classification, de-correlating data elements, discovering meaningful patterns and finding essential representation of correlated variables that form a manifold (e.g. the manifold of galaxies), tagging astronomical images, multiscale analysis synchronized across all available wavelengths, denoising, etc. The second part of this paper is dedicated to a new, active area of image processing: image inpainting that consists of automated methods for filling in missing or damaged regions in images. Inpainting has multiple astronomical applications including restoring images corrupted by instrument artifacts, removing undesirable objects like bright stars and their halos, sky estimating, and pre-processing for the Fourier or wavelet transforms. Applications of high-dimensional data reduction and mitigation of instrument artifacts are demonstrated on images taken by the Spitzer Space Telescope.

1. High-Dimensional Data Reduction: Approach and Results

Dimensionality reduction consists of methods for finding lower-dimensional representation of high-dimensional data by constructing a set of basis functions that capture patterns intrinsic to a particular state space. Examples of high-dimensional data include, but are not limited to, multiparametric data sets (e.g. the manifold of galaxies: Brosche 1973; Djorgovski & Davis 1987), multitemporal, multispectral and hyperspectral data sets and images. Dimension reduction facilitates data analysis, data representation and retrieval and improves statistical inference. All these tasks are crucial to multiwavelength astronomy, archival research, large-scale digital sky surveys and temporal astronomy. A cartoon of an elephant and blind men in Figure 1 illustrates the current situation in analysis of high-dimensional data in astronomy. We approach the problem more "holistically" by considering multidimensional data sets and images as manifolds or combinatorial graphs. First we develop algorithms for constructing variational splines that minimize certain Sobolev norms introduced in terms of the combinatorial Laplace operator (Pesenson 2005, 2006, 2008a; Pesenson & Pesenson 2008, 2009). As we have previously shown, variational splines allow approximate reconstruction (with any degree of accuracy) of the eigenfunctions and eigenvalues of the corresponding Laplace operator. This approach allows us to reconstruct low eigenvalues and corresponding eigenfunctions by using only a small portion of the graph. This result is a generalization of the Shannon sampling theorem. Another advantage of our splines is that they have very strong localization (similar to Euclidean variational splines) in contrast to eigenfunctions, thus enabling local analysis on graphs. We use these splines and the first two or three eigenfunctions of the Laplace operator to reconstruct embedding of large data sets into two- or three-dimensional Euclidean space.

An example of data reduction is shown in Figures 2 and 3. An immediate simple application of dimension reduction would be creating a multiwavelength "quick-look" image that includes all essential information in a statistically justified way, and thus is much more accurate than a "quick-look" made by simple coadding with an ad hoc, heuristic weighting.

Figure 1.: "An elephant and blind men" as a metaphor for high-dimensional data and the traditional approach to it: looking at different 'wavelengths' independently.

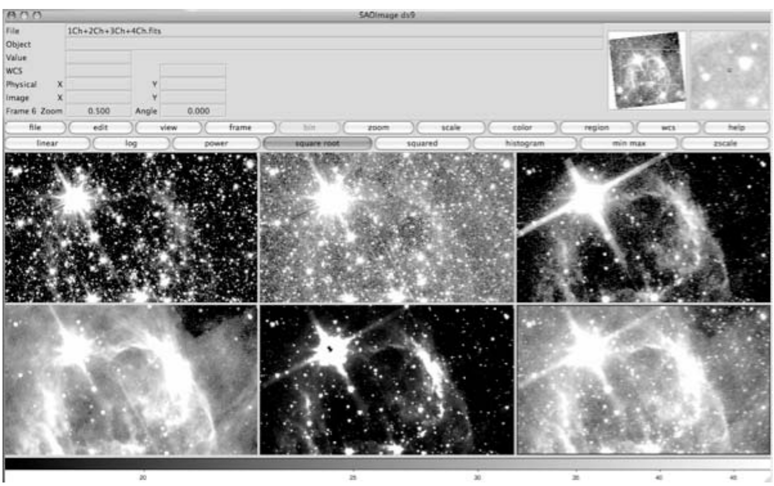

Figure 2.: SPITZER, IRAC. Upper row: channels 1-3. Lower row: channel 4, the result of applying dimension reduction, sum of the channels 1-4.

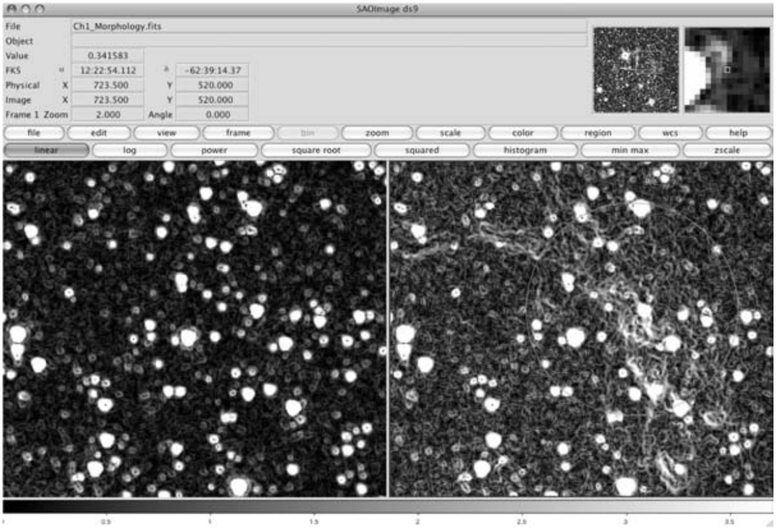

Figure 3.: SPITZER, IRAC, Left: channel 1, morphology unveiling following Pesenson et al. (2008). Right: morphology unveiling performed on the reduced image from Fig. 3. It demonstrates that a properly reduced image retains the important features from all channels. This is especially useful for dealing with large numbers of images.

2. Image Inpainting as Pre- and Post-processing

The motivation of image inpainting is to develop a tool for filling in missing or damaged regions in astronomical images in order to enable Fourier or wavelet transforms, mitigate artifacts, remove selected undesirable objects (bright stars), estimate sky, etc. (Bertalmio & Sapiro 2000; Tschumperle & Deriche 2005). Inpainting is done by using information from the surrounding area and propagating it from the surrounding area in a nonlinear way thus allowing it to restore structures (structural inpainting), texture (textural inpainting), or both. Some examples are demonstrated in Figures 4 and 5.

3. Conclusion

Dimension reduction methods provide an effective tool for multitemporal and multispectral data representation, analysis and visualization at the peta-scale (Graham 2009), thus becoming vital for the large sets of astronomical images obtained by LSST, Pan-STARRS, The Palomar Transient Factory and Palomar-Quest Survey (Tyson et al. 2007; Bailey et al. 2007; Law et al. 2009; Djorgovski et al. 2009). Dimensionally-reduced images also offer an enormous savings in storage space and database-transmission bandwidth for the user. However, the challenges now facing astronomy are not only about database sizes in themselves, but just as much about how intelligently one organizes and

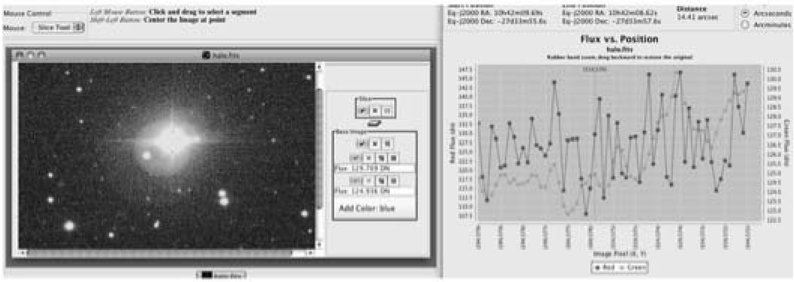

Figure 4.: ESO-LASILLA (courtesy of A. Grado, INAF-Osserv. Astr. di Capodimonte). Detecting a halo to be removed. Left: pre- and post-processed images overlaid. Right: flux-cut through the outmost left edge of the halo; the average level "outside" of the halo is lower for the post-processed image (green) than the average level inside.

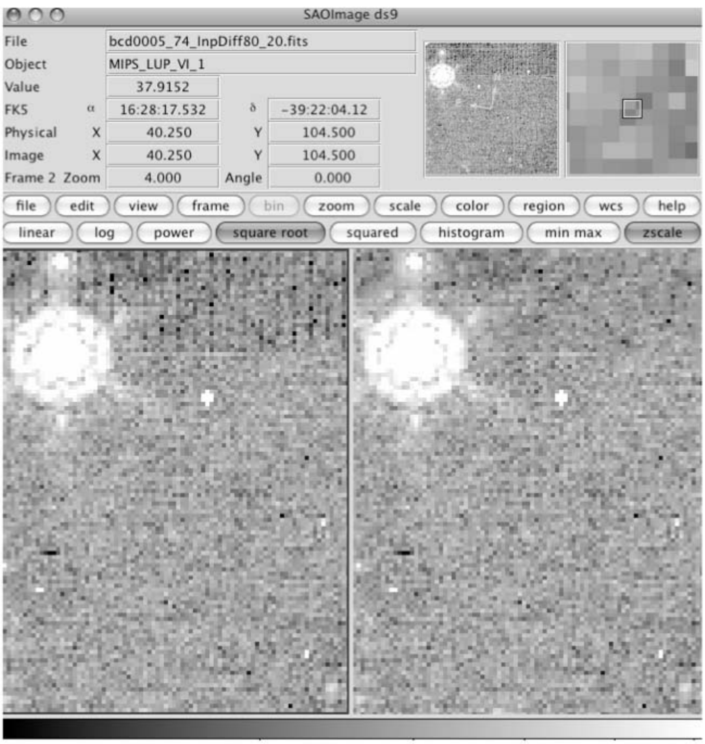

Figure 5.: Left: SPITZER, MIPS-24, instrument artifact - Jailbars. Right: the artifact removed by inpainting.

navigates through them. Dimension reduction methods play an important role in the emergence of the new era of semantic applications.

Acknowledgments. M. P. would like to thank I. Song, M. Egan, L. Ly, B. Hartley and T. Brooke for helpful discussions. This work was carried out with funding from NASA to the California Institute of Technology and the Jet Propulsion Laboratory and by the National Geospatial-Intelligence Agency University Research Initiative (NURI), grant HM1582-08-1-0019.

References

Bailey, S., et al. 2007, ApJ, 665, 1246
Bertalmio, M., Sapiro, G., Image Inpainting, SIGGRAPH 2000, 417
Brosche, P., 1973, A&A 23, 259
Djorgovski, S., Davis, M., 1987, ApJ, 313:59
Djorgovski, S., et al., 2009, 213th AAS Meeting, #427.02
Graham, M., 2009, in ASP Conf. Ser. 411, ADASS XVIII, ed. D.A. Bohlender, D. Durand & P. Dowler (San Francisco: ASP), 165
Law, N., et al., 2009, 213th AAS Meeting, #469.01
Pesenson, M., et al., 2009, 213th AAS Meeting, #427.04
Pesenson, M., Roby, W., McCollum, B. 2008, ApJ, v. 683:566
Pesenson, M., et al., 2007, in ASP Conf. Ser. 394, ADASS XVII, ed. Robert W. Argyle, Peter S. Bunclark & James R. Lewis (San Francisco: ASP), 97
Pesenson, M., et al., 2006, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco: ASP), 123
Pesenson, I., 2005, Journal of Approximation Theory, v.135, no. 2, 203220
Pesenson, I., 2006, Applied and Computational Harmonic Analysis, v. 21, no. 2, 230244
Pesenson, I., 2008a, SIAM Journal of Discrete Mathematics (accepted)
Pesenson, I., 2008b, Applied and Computational Harmonic Analysis (submitted)
Pesenson, I., Pesenson M., 2008, J. of Fourier Analysis and Applications (submitted)
Pesenson, I., Pesenson, M., 2009, Journal of Discrete Optimization (submitted)
Tschumperle, D., Deriche, R. 2005, IEEE Trans. PAMI., 27 (4), 1-12
Tyson, A., et al., 2007, in ASP Conf. Ser. 394, ADASS XVII, ed. Robert W. Argyle, Peter S. Bunclark & James R. Lewis (San Francisco: ASP), 107